Informatica

14 Oct, 2025

Data Quality Analytics Framework

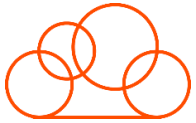
- Mayank Kumar Shrivastava, Senior Principal Solutions Architect, IPS
- Ding Han Tan, Senior Consultant, IPS

Where data & AI come to LIFE

Housekeeping Tips

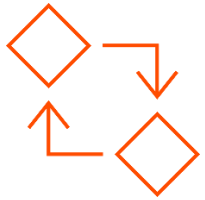
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the Q&A option and we will respond at the end of the presentation
- The webinar is being recorded and will be available on our Success Portal - where you can download the slide deck for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the post-webinar survey and provide your feedback and suggestions for upcoming topics.

Feature Rich Success Portal



Bootstrap trial and POC Customers

Enriched Customer Onboarding experience

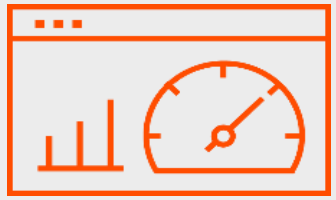


Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

More Information



Success Portal

https://success.informatica.com

Communities & Support

https://network.informatica.com

Documentation

https://docs.informatica.com

University

https://www.informatica.com/in/services-and-training/informatica-university.html

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Agenda

1. Use Cases
2. Solution Overview
3. Technical Architecture
4. Data Model
5. Solution High Level Implementation Steps
6. Sample Outputs and Dashboards
7. Demo
8. Business Outcomes

Use Cases

- Able to calculate weightage average Data Quality scores at various Business Assets such as Business Term, Business Area, Domain and Technical Assets like System, Table, Schema etc.

- Able to create complex dashboards and metrics showing key insights such as below:
  - Percentage of CDEs meeting DG requirements like Stakeholdership, Required Attributes grouped by Domain
  - Number of CDEs identified per Domain and grouped by Business Area
  - Data Quality Coverage on CDEs grouped by Domain
  - Data Quality score per DQ Dimension grouped by Legal Entities

- Able to perform trend analysis for Data Quality scores at various Business Assets and Technical assets.

Solution Overview

- Platform
  Intelligent Data Management Cloud (IDMC) platform.

- Data Source:
  Data Source can be any RDBMS or Files but, in our demo, we have used Databricks.

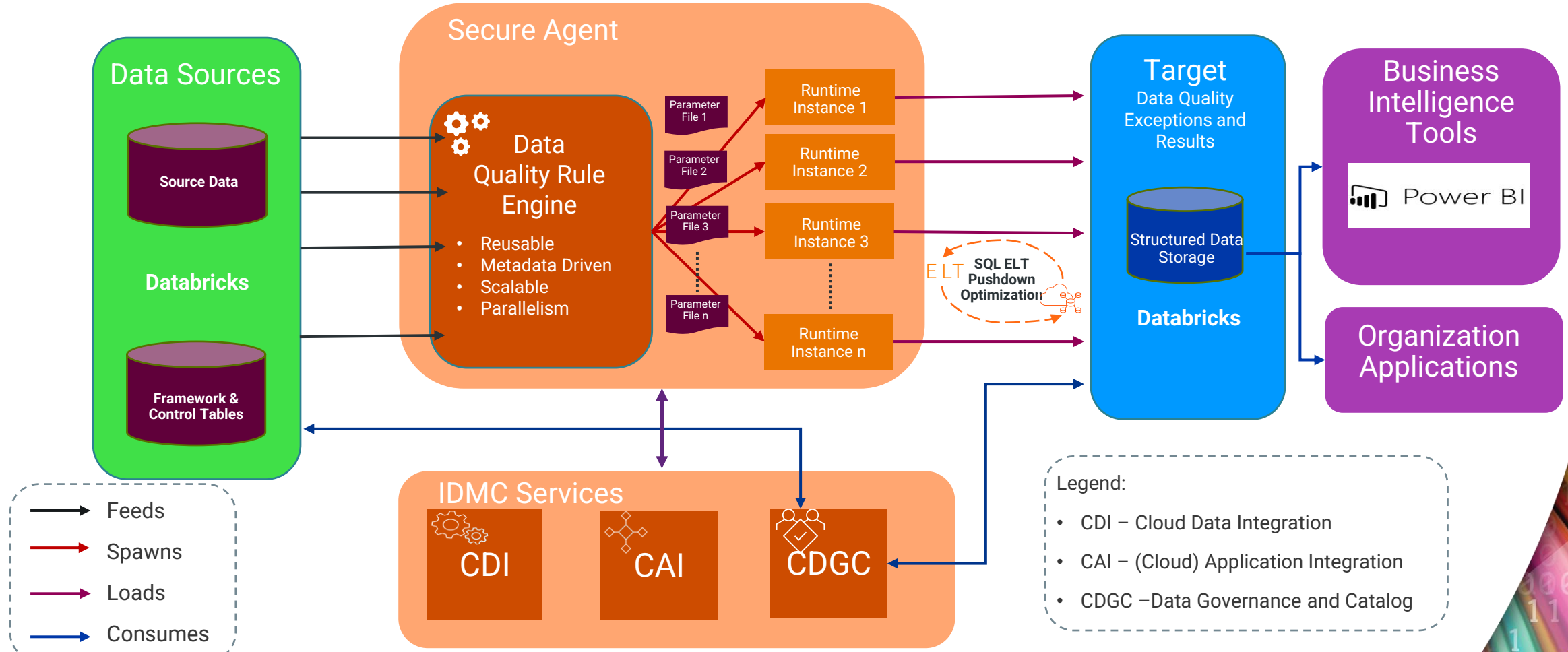
- Key IDMC Components:
  Core logic designed and orchestrated within the Informatica IDMC platform using:
  - ❑ Cloud Data Integration (CDI): Designs and executes primary DQ transformations and rules.
  - ❑ Cloud Application Integration (CAI): Manages end-to-end orchestration, job control, and API-triggered jobs.
  - ❑ Cloud Data Governance & Catalog (CDGC): Publishes and tracks DQ scores, making them visible enterprise-wide.

- Execution Model:
  Template-driven and parameterized approach standardizes DQ jobs, enabling rapid onboarding of new datasets and rules.

- Target:
  DQ results, exceptions and other dimension tables can be stored in RDMS, Cloud Data Warehouses, Cloud Data Lake, in our demo, we have used Databricks. These then can be exposed to Data Visualization tools such as Power BI, Tableau.

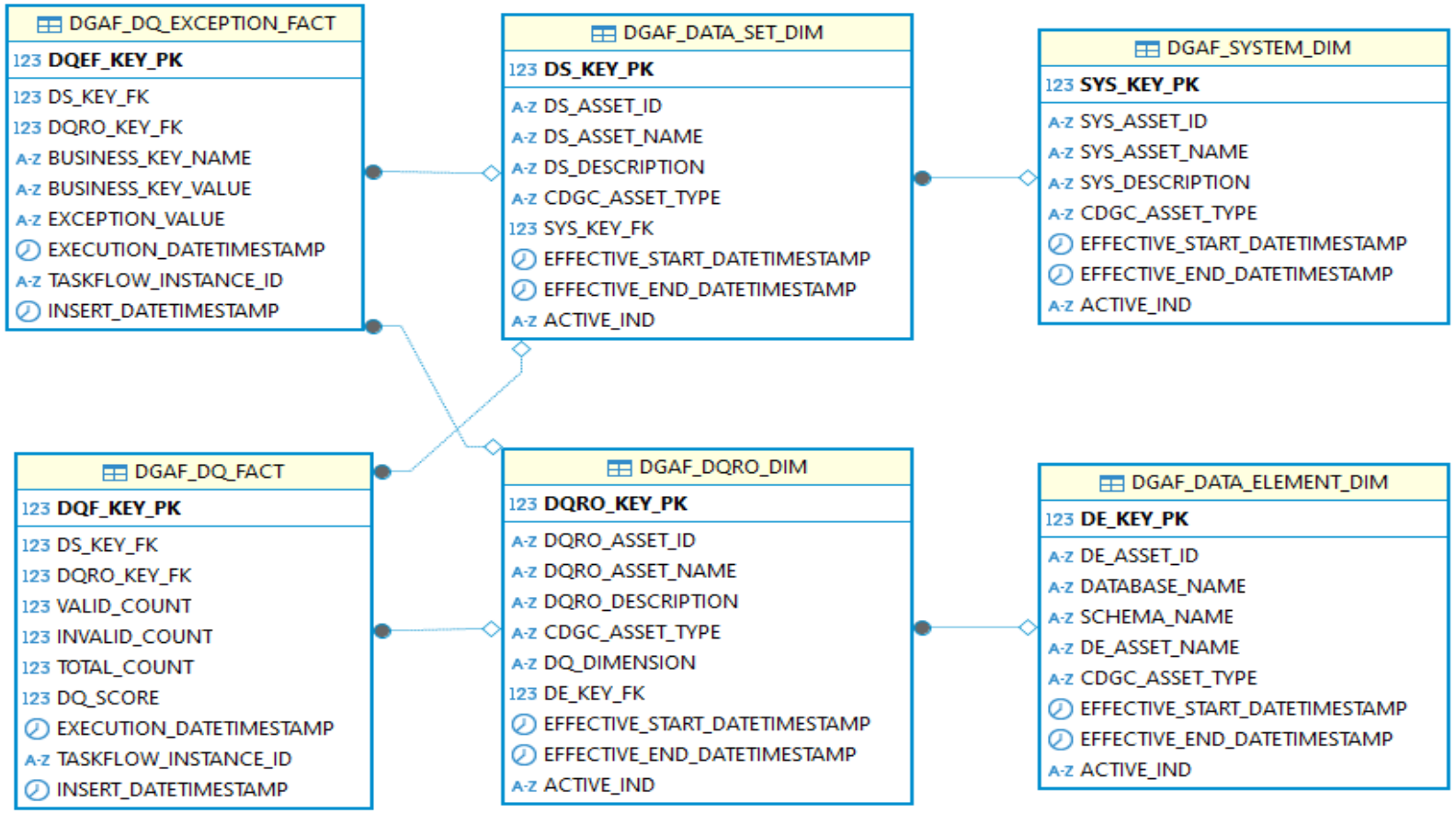
- Uploading DQ scores to CDGC
  As part of generating the DQ results and exceptions, there will be a calculation of DQ scores which will be pushed to CDGC.

Data Quality Analytics Solution Architecture



Informatica

Data Model



Solution High Level Implementation Steps

- Step-1: Capture Data Quality rules and its related business assets like System, Data Set, Data Element, Glossary etc. in Informatica Cloud Data Governance & Catalog (CDGC).

- Step-2: Create reusable Data Quality Rules as Mapplets in Informatica Cloud Data Integration.

- Step-3: Create parameterized Mapping in Informatica Cloud Data Integration which read data from source table/file and generates Data Quality Result and Exceptions in target like Databricks.

- Step-4: Create a mapping that reads Data Quality results table data and generate JSON file containing required details like Execution Time, DQ Score, Exception Count, Total Count, Identity etc. to push DQ Results to Informatica CDGC.

Solution High Level Implementation Steps

- Step-5: Create the logic to read JSON file data and make an API call to Informatica CDGC to push the DQ Results.

- Step-6:. Create Mappings to fetch required CDGC Business and Technical assets and write to the data model already setup in Databricks.

- Step-7: Orchestrate all created Mappings and other logics like Notifications in Informatica Cloud Application Integration and publish it.

- Step-8: Execute the job either via IDMC scheduler or third-party applications such Python based code, Control-M application.

Sample Output

DQ Result

Table

	sys_asset_name	ds_asset_name	ds_description	dqro_asset_name	dqro_description	de_asset_name
1	Databricks	Address	Contains Address Information <...	Address_ZIP_CODE_Validity_...	Validity check on ZIP_CODE (PDE)...	ZIP_CODE
2	Databricks	Address	Contains Address Information <...	Address_ADDRESS_Complete...	Completeness check on ADDRES...	ADDRESS

dq_dimension	valid_count	invalid_count	total_count	dq_score	execution_datetimestamp	taskflow_instance_id
Validity	1752	799	2551	68.68	2025-10-02T20:37:14.000+00:00	1158419810702684160
Completeness	2551	0	2551	100.00	2025-10-02T20:37:14.000+00:00	1158419810702684160

Table

	sys_asset_name	ds_asset_name	ds_description	dqro_asset_name	dqro_description	de_asset_name
1	Databricks	Party	Contains Party Information ...	Party_EMAIL_Validity_Check	Validity check on EMAIL (PDE) in Party Table	EMAIL
2	Databricks	Party	Contains Party Information ...	Party_NAME_Completeness_Check	Completeness check on NAME(PDE) in Party Table	NAME

dq_dimension	valid_count	invalid_count	total_count	dq_score	execution_datetimestamp	taskflow_instance_id
Validity	683	73	756	90.34	2025-10-02T20:37:15.000+00:00	1158419820127289344
Completeness	746	10	756	98.68	2025-10-02T20:37:15.000+00:00	1158419820127289344

Sample Output

DQ Exceptions

Table +

	sys_asset_name	ds_asset_name	ds_description	dqro_asset_name	dqro_description	de_asset_name
1	Databricks	Address	> <p>Contains Address...	Address_ZIP_CODE_Validity_Check	Validity check on ZIP_CODE (PDE) in Address Table	ZIP_CODE
2	Databricks	Address	> <p>Contains Address...	Address_ZIP_CODE_Validity_Check	Validity check on ZIP_CODE (PDE) in Address Table	ZIP_CODE
3	Databricks	Address	> <p>Contains Address...	Address_ZIP_CODE_Validity_Check	Validity check on ZIP_CODE (PDE) in Address Table	ZIP_CODE
4	Databricks	Address	> <p>Contains Address...	Address_ZIP_CODE_Validity_Check	Validity check on ZIP_CODE (PDE) in Address Table	ZIP_CODE
5	Databricks	Address	> <p>Contains Address...	Address_ZIP_CODE_Validity_Check	Validity check on ZIP_CODE (PDE) in Address Table	ZIP_CODE

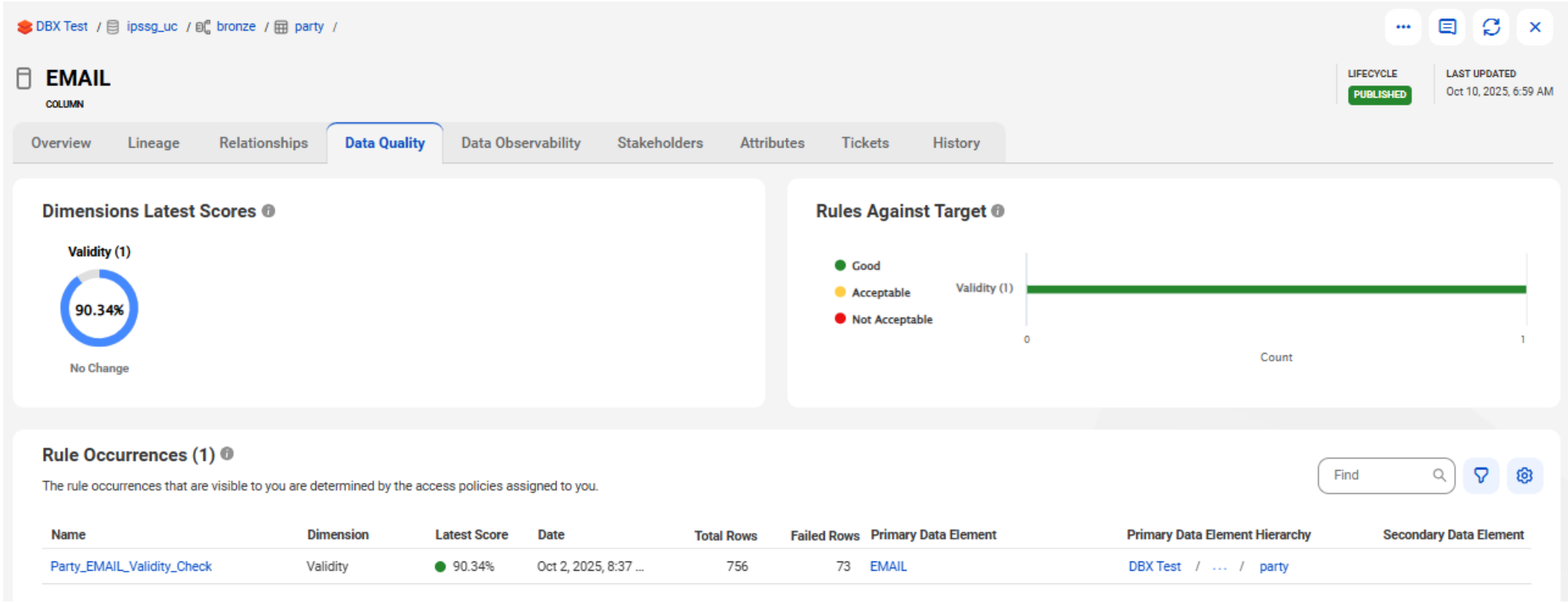
dq_dimension	business_key_name	business_key_value	exception_value	execution_datetimestamp	taskflow_instance_id
Validity	CUST_ID	1080	L5R 6R4	2025-10-02T20:37:14.000+00:00	1158419810702684160
Validity	CUST_ID	1073	L2X 0A0	2025-10-02T20:37:14.000+00:00	1158419810702684160
Validity	CUST_ID	1077	L5P 0A0	2025-10-02T20:37:14.000+00:00	1158419810702684160
Validity	CUST_ID	1088	M9R 3L6	2025-10-02T20:37:14.000+00:00	1158419810702684160
Validity	CUST_ID	1089	H7L 4T7	2025-10-02T20:37:14.000+00:00	1158419810702684160

Table +

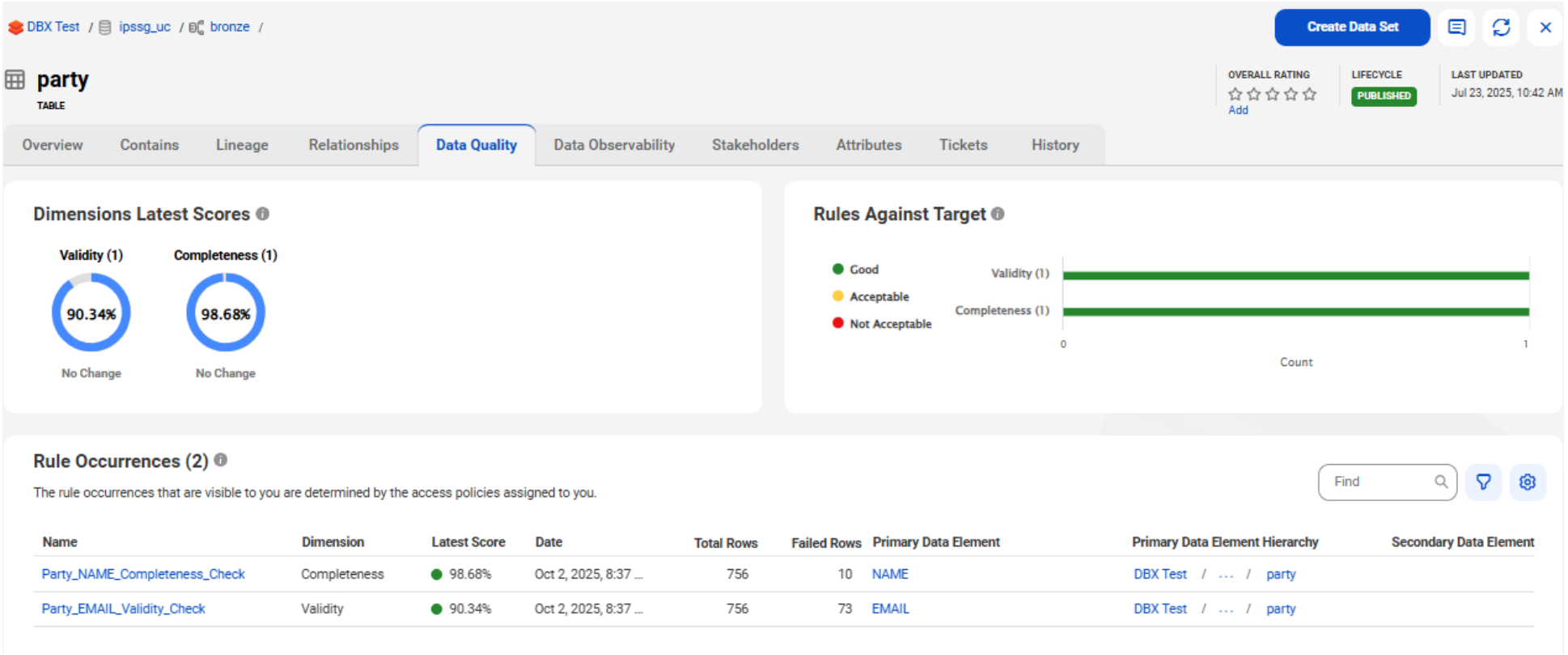
	sys_asset_name	ds_asset_name	ds_description	dqro_asset_name	dqro_description	de_asset_name
1	Databricks	Party	> <p>Contains Party Information <...	Party_EMAIL_Validity_Check	Validity check on EMAIL (PDE) in Party Table	EMAIL
2	Databricks	Party	> <p>Contains Party Information <...	Party_EMAIL_Validity_Check	Validity check on EMAIL (PDE) in Party Table	EMAIL
3	Databricks	Party	> <p>Contains Party Information <...	Party_EMAIL_Validity_Check	Validity check on EMAIL (PDE) in Party Table	EMAIL
4	Databricks	Party	> <p>Contains Party Information <...	Party_EMAIL_Validity_Check	Validity check on EMAIL (PDE) in Party Table	EMAIL
5	Databricks	Party	> <p>Contains Party Information <...	Party_EMAIL_Validity_Check	Validity check on EMAIL (PDE) in Party Table	EMAIL

dq_dimension	business_key_name	business_key_value	exception_value	execution_datetimestamp	taskflow_instance_id
Validity	PARTY_ID	2774100	super stop & shop@stopandshop.com	2025-10-02T20:37:15.000+00:00	1158419820127289344
Validity	PARTY_ID	3339000	key food@keyfoodstores.com	2025-10-02T20:37:15.000+00:00	1158419820127289344
Validity	PARTY_ID	3374700	key food@keyfoodstores.com	2025-10-02T20:37:15.000+00:00	1158419820127289344
Validity	PARTY_ID	3355800	key food@keyfoodstores.com	2025-10-02T20:37:15.000+00:00	1158419820127289344
Validity	PARTY_ID	3383100	key food@keyfoodstores.com	2025-10-02T20:37:15.000+00:00	1158419820127289344

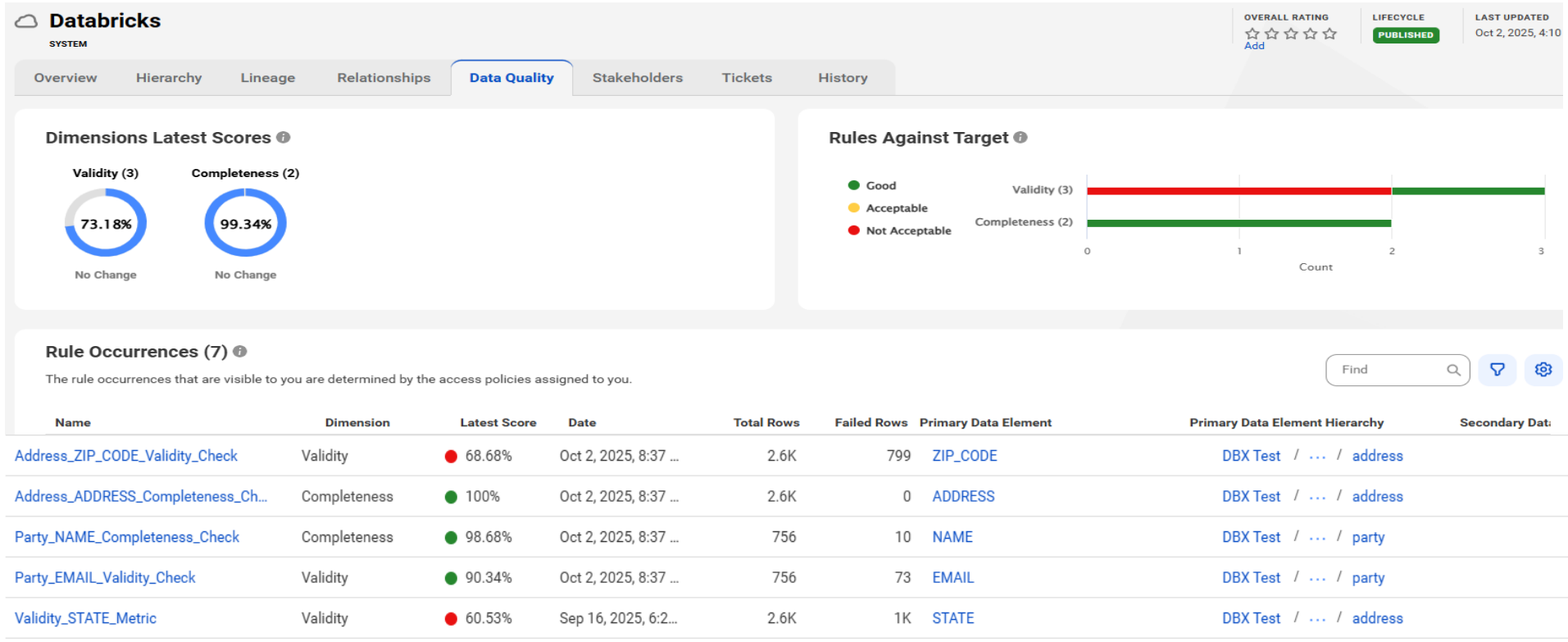
Sample CDGC DQ Dashboard – Data Element



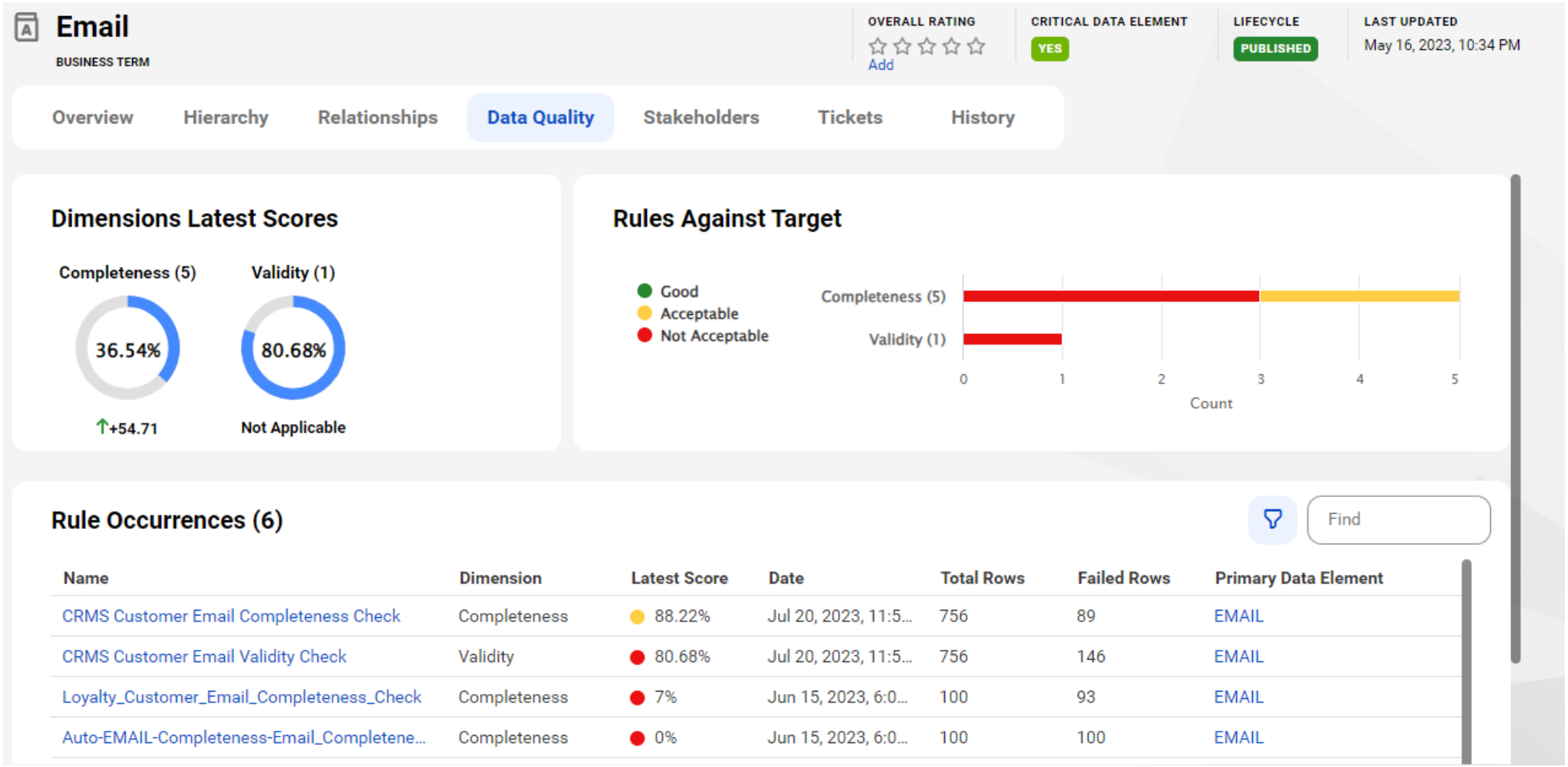
Sample CDGC DQ Dashboard – Table



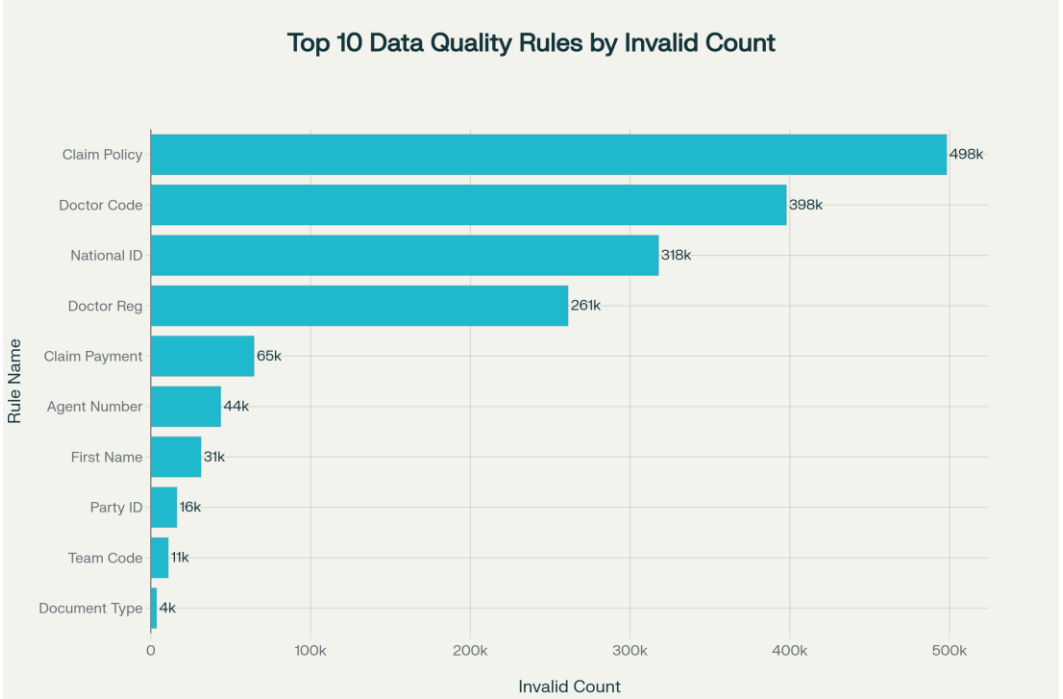
Sample CDGC DQ Dashboard – System



Sample CDGC DQ Dashboard – Glossary



Sample BI Dashboards



Informatica

DEMO

Where data & AI come to LIFE

Business Outcomes

AI-Driven Custom Reporting & Insights
Integrate AI models to enable flexible, user-driven dashboards and natural language querying, allowing users to ask questions about data quality and receive tailored insights dynamically.

Advanced Predictive Analytics
Leverage machine learning to proactively identify potential data quality issues and suggest targeted remediation based on historical trends and anomalies.

Deeper Business Context Linking
Enhance alignment of technical data quality findings with business impact and operational workflows for prioritized, actionable governance.

Where data & AI come to

Informatica